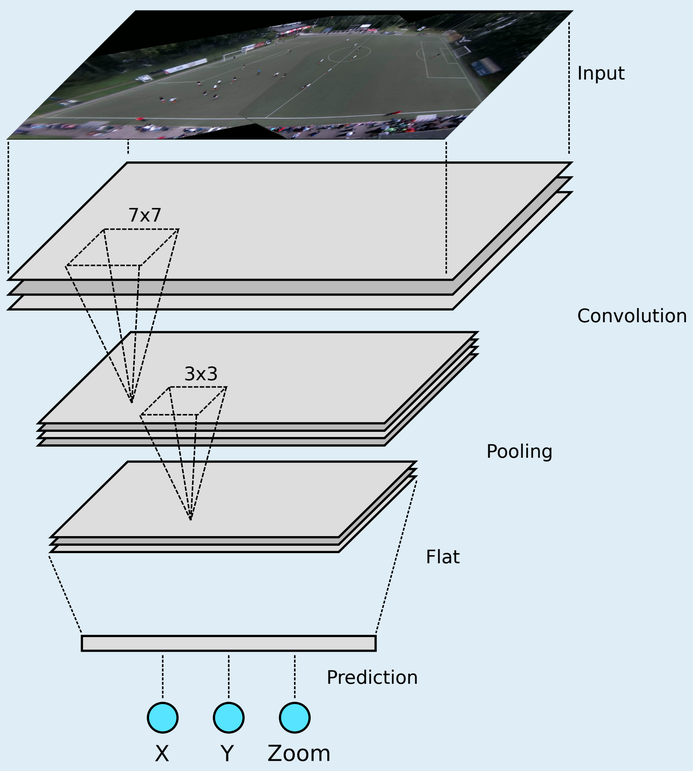
Automated Soccer Scene Tracking Using Deep Neural Networks

C. Bodenstein*, M.Goetz*, M. Riedel*

* High Productivity Data Processing Research Group, Jülich Supercomputing Center (JSC)



Construction of an automated pipeline for the broadcast of football games

- In Germany: Up to 80k football games each week
- Most matches will never be recorded e.g. amateurs
- TV camera systems and cinematographer are expensive
- Simple object tracking—i.e. the ball—is not sufficient for specific game situations like e.g. corners
- Learn the scene tracking using Deep Neural Networks
- Goal: determine the point of interest coordinates and camera zoom for each frame

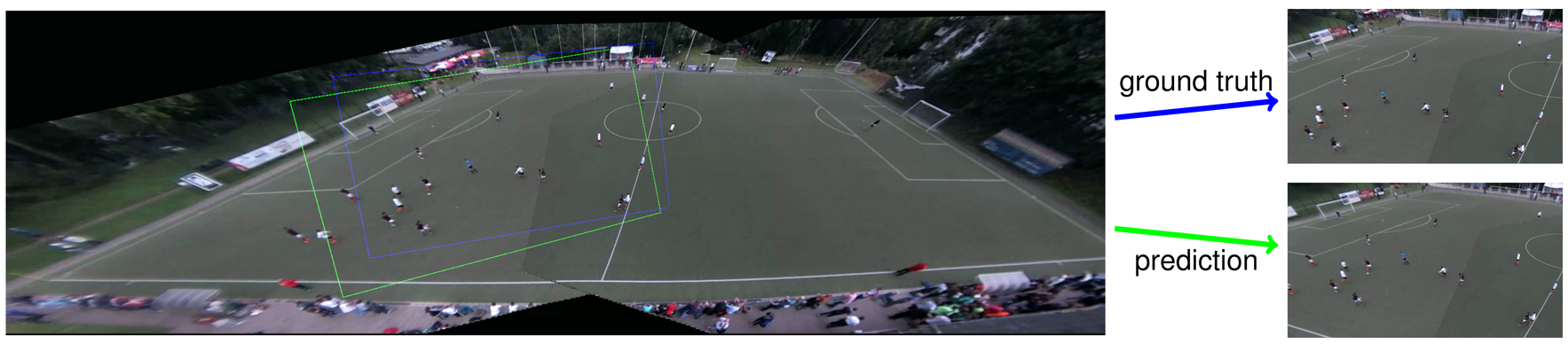


Figure: Panorama of the entire football field. The tetragons represent the focused areas by the cameraman. ● Manually captured ground truth, ● Prediction made by Deep Neural Network

Data Source

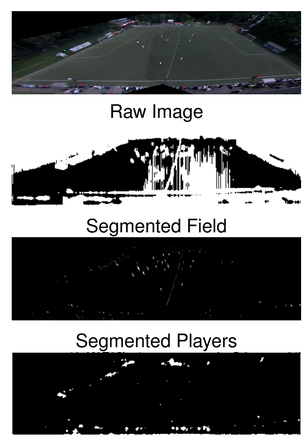
- Capture the entire field with multiple static cameras—two prototypes with 2 or 5 images respectively
- Stitch images to singular panorama
- Labeled focus x, y point and zoom
- Currently 30 labeled football games
- ∼0.5 TB MJPEG compressed
- ∼1.500.000 Frames
- High resolution images in 2017
- More at www.soccerwatch.tv

Why Deep Learning?

- Convolutional Neural Networks (CNNs) are state-of-the-art in image classification and object tracking [1]
- Layers abstract different things: the audience, players, the ball, etc.
- Recurrent Neural Networks retain time information from sequentially analyzed frames [2]
- Popularity resulted in highly optimized tools that use GPUs

The Learning Outcome

- Input: DNN gets a sequence of frames
- Output: Three output neurons describing x and y position of the camera focus plus the additional zoom
- Prediction close to the ground truth and appears natural
- A look into the convolutions can show learned features



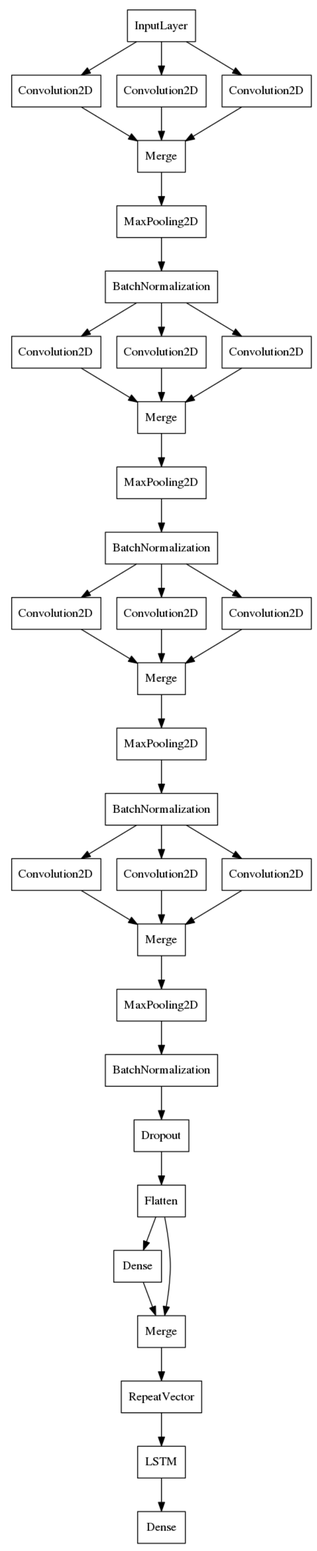
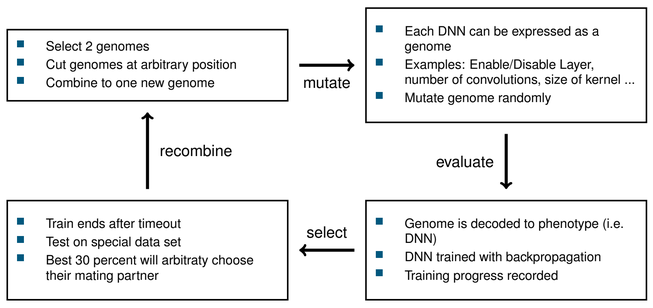


Figure: Network architecture

Genetic Optimization of the Deep Neural Networks



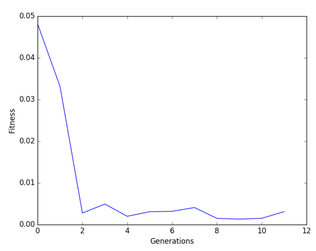


Figure: Fitness over generations (lower is better)

- Genetic optimization of Network Architecture on Apache Spark Cluster
- 20 different DNN ”individuals” in parallel
- Selection after three hours of learning

Future Work

- Training on new, high-quality data
- Evaluation of the usage of RNNs
- Reduction of the Network’s complexity to allow real time performance—5 FPS is sufficient
- Smooth the camera motions in a post processing step

References

[1] Steven J Nowlan and John C Platt. A convolutional neural network hand tracker. *Advances in Neural Information Processing Systems*, pages 901–908, 1995. [2] Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, and George Toderici. Beyond short snippets: Deep networks for video classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4694–4702, 2015.

Contact: c.bodenstein | m.goetz | m.riedel@fz-juelich.de **- Website:** www.fz-juelich.de **Camera system:** info@soccerwatch.tv